Markov Random Field Based Pruning and Learning Based Rescoring for Object Detection

Naoto Inoue, Ryosuke Furuta, Toshihiko Yamasaki, Kiyoharu Aizawa

Dept. Information and Communication Engineering, The University of Tokyo

**Background**

“Context” in object detection
- information out of each candidate window
- co-occurrence of objects, spatial layout, scale, background, scene of image...

R-CNN based methods do not consider context.

**Proposed Method**

1. **Pruning by MRF**
   - Decide whether each window should be pruned ($w = 1$) or not ($w = 0$).
   - Better than just setting threshold.

2. **Rescoring by SVM**
   - Predict whether each window is correct by SVM.
   - New score $s_w$ is calculated based on decision value $d_w$.

**Experimental Results**

**Evaluation**

Dataset: VOC2007 and MSCOCO.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>VOC2007-test</th>
<th>MSCOCO-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation metric</td>
<td>mAP [%]</td>
<td>F1 [%]</td>
</tr>
<tr>
<td>Baseline (Fast R-CNN)</td>
<td>66.9</td>
<td>3.5</td>
</tr>
<tr>
<td>+ Tree Context [Choi+, 2012]</td>
<td>60.9 (6.0)</td>
<td>3.5 (+0.0)</td>
</tr>
<tr>
<td>+ HOOD [Cao+, 2015]</td>
<td>57.9 (-0.9)</td>
<td>67.2 (+63.7)</td>
</tr>
<tr>
<td>+ threshold</td>
<td>66.5 (-0.4)</td>
<td>24.4 (+20.9)</td>
</tr>
<tr>
<td>+ pruning</td>
<td>66.5 (-0.4)</td>
<td>26.2 (+22.7)</td>
</tr>
<tr>
<td>+ pruning + rescoring</td>
<td>67.3 (+0.4)</td>
<td>26.2 (+22.7)</td>
</tr>
</tbody>
</table>

**Conclusion**

- Reducing and rescoring candidate windows by considering contextual model.
- Fast R-CNN detectors are improved by +0.4% on mAP and +22.7% on F1 in VOC2007-test.
+0.7% on mAP and +2.6% on F1 in MSCOCO-val.
- Applications to structured retrieval are also presented.